

Identifying and Characterizing a “Consumer Medical Vocabulary”
Doctoral Dissertation Abstract

Anthony Y. Tse, Doctor of Philosophy, 2003

Directed by Professor Dagobert Soergel
College of Information Studies, University of Maryland at College Park

This dissertation explores procedures for collecting and analyzing terms used by non-professionals during discourse on medical topics in order to construct a Consumer Medical Vocabulary and ultimately to learn more about laypersons’ understanding of medical concepts. It describes methods for term extraction and vocabulary generation from documents representing two discourse groups: (1) participants in Web-based health discussion forums (“consumers”) and (2) writers of newspaper and magazine articles, informational pamphlets for patients, and articles from consumer health Web sites (“mediators”).

Over 100,000 forms were extracted from documents in the consumer and mediator corpora manually by laypersons. The forms were mapped to concepts from the National Library of Medicine’s (NLM) Unified Medical Language System[®] (UMLS[®]) Metathesaurus[®], using a semi-automated process. The forms and concepts used by consumers (CMV) and mediators (MMV) were analyzed and compared with medical terms used by healthcare professionals (PMV), using descriptive methods. The analysis compared vocabulary characteristics in 15 medical subdomains, such as “Diseases” and “Drugs.”

The preliminary findings suggest that the terms used by the non-professional discourse groups studied (CMV and MMV) differed from those used by professionals (PMV) more by form than by concept. Furthermore, categories of forms commonly used by consumers (“consensus” forms) were observed.

This exploratory work contributes to a number of areas. After further validation, patterns in forms used by non-professionals may be used for developing automated term extraction processes. The corpus-based term identification, review, and processing procedure used in this study may be adapted to study the use of terminology among discourse groups in other domains. Additional study of consensus forms may help in the development of a consumer health thesaurus to facilitate lay access to professional terms.